DOCUMENT RESUME

ED 193 678 CS 205 947

AUTHOR Mead, Nancy A.
TITLE Assessing Listening Ability: Relationships with Verbal Ability and Racial/Ethnic Bias.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
REPORT NO NAEP-08-SL-53
PUB DATE Nov 78
NOTE 14p.: Paper presented at the Annual Meeting of the Speech Communication Association (64th, Minneapolis, MN, November 2-5, 1978).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement: Adolescents: *Educational Research: *Listening Comprehension Tests: *Listening Skills: *Racial Bias: Secondary Education: *Test Bias: Test Construction: *Verbal Ability

ABSTRACT

Two studies were conducted to assess the listening ability of 17-year-old students with particular interest in the relationship among listening ability, verbal ability, and racial/ethnic bias in the test items. The first study, a National Assessment of Educational Progress and Speech Communication Association pilot project, indicated that there was potential item bias based on minority ethnic group response. However, no real data on the students' verbal ability was available to substantiate possible explanations. In the second study, two packages of listening items, each approximately 35 minutes in length were developed, and a 25-item vocabulary test was administered to estimate verbal ability. Recording of the listening items was made by individuals using network English. The results indicated that there was a positive correlation between listening ability and verbal ability, suggesting that the differing responses of minority and nonminority students in the first study could be attributed to differing levels of verbal ability, rather than item bias. (MKM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED193678

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ASSESSING LISTENING ABILITY: RELATIONSHIPS WITH VERBAL ABILITY AND RACIAL/ETHNIC BIAS

No. 08-SL-53

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
T. Pratt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by

Nancy A. Mead

National Assessment of Educational Progress

Denver, Colorado

Presented at the Annual Convention of the Speech Communication Association

Minneapolis, Minnesota

November 1978

CS205947

Introduction

One of the important aspects of communication competence is listening ability. This is a complex set of skills which require both careful recall and critical analysis of spoken messages. In a society which is barraged with political speeches, commercials, lectures and commands, it seems critical that listening ability be an important focus of communication research, instruction and assessment. The present paper discusses attempts to assess the listening ability of seventeen-year-olds. It describes two studies which developed and field tested measures in this area and some of the problems and conclusions which were reached concerning the assessment of listening ability. Of particular interest was the relationships among listening ability, verbal ability and racial/ethnic bias in test items.

Speaking and Listening Assessment Pilot Project

In June of 1976 the National Assessment of Educational Progress (NAEP) and the Speech Communication Association (SCA) initiated a pilot study to test the feasibility of assessing speaking and listening skills (Mead, 1977a, 1977b). The products of this effort were intended for use in the National Assessment of Educational Progress, a national survey of student achievement with respect to important educational objectives, funded by the National Center for Education Statistics.

There are some important differences between National Assessment and standardized achievement testing programs. The items developed by NAEP measure specific objectives which are considered important by educators and content specialists. They do not constitute a test per se. The items

are used to describe the accomplishments of nationally representative groups of students. They are not used to differentiate levels of ability among individual students. Nevertheless, the task of developing listening assessment items was similar to standardized test development in that it involved defining the domain of listening ability and constructing items which measured that domain.

As a part of this study, a pool of fifty-six listening items was developed. The items were packaged into four test booklets, each representing approximately fifteen minutes of testing. The items were field tested in four sites which represented a variety of regions of the country, size and type of cities, and racial and ethnic populations. An average of 140 students responded to each set of items.

The items were analyzed using typical item analysis statistics. Item difficulty was indicated by the percent of students choosing each option. Item discrimination was indicated by the point biserial correlation between individuals choosing an option and their total test scores. In addition, the responses to each option were correlated with an external criterion which reflected the classification of the students as minority or non-minority. This added information allowed reviewers to identify items which received significantly different responses by minority and nonminority students.

It is important to emphasize the purpose of adding the external criterion which reflected the racial/ethnic background of students to the information base. The aim of this strategy was not to eliminate all items which differentiated between minority and nonminority students. It is possible that there are real differences between these two groups with

respect to listening ability. The information was used to identify items which might be discriminating between minority and nonminority students for reasons other than listening skill. For example, an item might receive different responses because of the varying backgrounds, experiences, values or language styles of minority and nonminority students. We considered these factors to be extraneous to listening ability.

The results of field testing indicated no problems with respect to selecting items with appropriate difficulty level and discrimination power. Guidelines had been established for selecting items within the difficulty range of forty percent to eighty percent correct responses with an average of sixty percent (Stanley and Hopkins, 1972) and with a discrimination level of at least .30 (Harris, 1968). Because the purpose of National Assessment is not to build a test but to select items which measure specific objectives, these guidelines were merely suggestive and not crucial. Practically all of the items in the pool met the discrimination requirement and only about twenty percent of the items fell outside the proposed difficulty range.

The surprising result from tryouts was that the listening items, unlike the items for the other areas of communication competence which were field tested at the same time, showed a high number of significant point biserial correlations between responses and minority/nonminority status. Approximately one-half of the listening items demonstrated this characteristic. It must be emphasized that a significant correlation between the responses and minority/nonminority status (a relationship significantly different from zero) was not considered tantamount to item bias. There were a couple of reasons for reviewing the data cautiously. First, the

tryout sites included two all minority schools. This made it possible that the distributions might include a concentration of minority students within a single option because of some unusual responses by the students in these schools. Secondly, a great number of correlations were reviewed, one for each foil of each item. Among these, there were bound to be some significant relationships due to chance (one out of twenty).

A significant correlation was considered a signal for further review. In some cases, the critique indicated possible sources of item bias, such as a typically white speech pattern in a listening stimuli which presented a persuasive conversation between two friends, and the item was dropped. In other cases, however, the review could not detect any problems and the item was retained. As indicated earlier, a significant correlation was not considered to be synonymous with item bias. However, the frequency of this characteristic and the marked difference between this set of items and the other sets of items (informing speaking, controlling speaking, ritualizing and sharing feelings) suggested a special problem.

A panel of speech communication experts reviewed the listening items and selected approximately one-half for use in the assessment. About one-third of the selected items reflected significantly different responses by minority and nonminority students. The consultants identified very few specific aspects of the listening items which they felt were indicative of item bias, such as the type of situation presented, the speech style used, or the values implied. However, they speculated a number of general characteristics of the items which might have tapped factors which were extraneous to measuring listening ability. These problems included:

1) the vocabulary level of the listening stimuli;
2) the length of the formal speeches;
3) the interest level of the listening stimuli;
4) the accent and rate of speech of the speakers on the stimulus
5 tapes; and
5) the level of disruption in the classrooms.

They hypothesized that minority students might have less specialized vocabulary knowledge; a lower tolerance for long, boring materials; and less experience listening to the accents and rate of white speakers. Furthermore, minority students might tend to be concentrated in schools where there were more disruptions in the classrooms and nearby environment.

An additional factor which might explain the results is varying levels of verbal ability of the minority and nonminority students in the tryout groups. If listening ability overlaps with verbal ability, as previous research indicates, it is possible that the results might be partially explained in terms of different levels of verbal ability. The field testing did not collect information about the verbal ability of the students. It is possible that the minority students selected for tryouts reflected an overall lower level of verbal ability than the nonminority students.

The outcome of the tryout phase of the pilot listening assessment was the identification of a problem, potential item bias, and no real data to substantiate or further elaborate the situation. A number of explanations of the results were proposed. However, these explanations were based on speculation and not on empirical evidence. The problem of minority bias had not been clearly articulated in past listening assessment efforts. The results of the NAEP/SCA pilot project suggested a clear need for further development and research.

Listening Assessment Follow-Up Study

A listening assessment follow-up study was conducted to alleviate the problems identified by the reviewers of the previous study and to explore more definitively the relationships among listening ability, verbal ability, and racial/ethnic bias. Two packages of listening items, each approximately 35 minutes in length, were developed following the suggestions of the reviewers. Some materials from the earlier study were rewritten and others were added. The final set of materials included five informal informative listening situations--a newscast, a public service announcement, a weather report, a telephone call and a public address announcement--and two informal persuasive listening situations--a paid political announcement and a commercial. Also included were two speeches, one about pollution and another about alien workers, both with persuasive elements. One interactive situation was included, an interview between an employer and a teenage job seeker. These materials were kept purposefully short, all within the range of thirty seconds to three minutes. We used the Dale-Chall and Fry readability formulas to measure the difficulty of the material and found that the readability of the materials ranged from fifth grade to college level.

Questions about the listening stimuli were developed which reflected five types of listening tasks:

1) to be able to recall significant details;
2) to be able to comprehend the main idea;
3) to be able to draw inferences about the information (e.g., relationships, implications);
4) to be able to make judgments concerning the speaker (e.g., intent, attitudes); and
5) to be able to make judgments concerning the information (e.g., types of evidence, logic of arguments).

Recordings of the materials were made by individuals using network English.

The final set of listening stimuli and questions, we believe, responded to the criticisms of reviewers. Each of the two packages of materials was field tested with four classrooms of eleventh graders in four sites, representing four parts of the country and a variety of sizes and types of city. There were a total of 84 respondents for package one and 86 for package two.

In order to explore the relationship among listening ability, verbal ability and racial/ethnic bias, additional data were collected. All students responded to a 25-item vocabulary test, a subtest of the Lorge-Thorndike Intelligence Tests (Lorge, Thorndike and Hagen, 1966). This was used to estimate the verbal ability of the students. The respondents to package one included six percent minority students and the respondents to package two included fourteen percent minority students. The minority representation was lower than expected. Nevertheless, the results still indicated some interesting relationships.

The results of the field test showed that the two packages of materials were performing similarly. The mean listening score for the first package was 59.5 percent correct with a standard deviation of 4.86. The mean score for the second package was 55.9 percent correct with a standard deviation of 5.51. The students who responded to the two sets of items were similar in verbal ability. The students who took package one had a mean vocabulary score of 11.77 and the students who took package two had a mean vocabulary score of 10.66. An interesting contrast in the two groups was that the mean vocabulary score of the minority students responding to package one was 12.8 and the mean vocabulary score of the minority students responding to package two was 6.7. Thus, the minority students responding to package one were above the average for their group and the minority students

responding to package two were well below the average of their group.

As might be expected in a field test, the individual item statistics indicated that some items were clearly not operating well. With respect to item difficulty, ten items in package one and nine in package two failed to fall within the desired range of forty percent to eighty percent correct (Stanley and Hopkins, 1972). With respect to item discrimination, sixteen items in package one and six in package two failed to meet the desired item-total correlation of .30 (Harris, 1968). Unlike the results in the first study, the items in these two packages did not show significantly different responses for minority and nonminority students. Only one item displayed this characteristic. It seemed clear that from the item pool it would be possible to pick adequate stimuli and items to use in one finalized package.

Even with the problems with individual items cited above, the items as a whole appeared to be operating quite well. The reliability of items in package one was .78 and in package two was .73 (using Hoyt's measure of internal consistency). It, therefore, seemed appropriate to proceed to look at the interrelationships among listening ability, verbal ability and racial/ethnic bias. It seemed curious that the problem of possible racial/ethnic bias which had been so prevalent in the first study was not evident in the second set of items. Indeed, steps had been taken to alleviate the possible problems of bias by controlling the vocabulary level, decreasing length, increasing interest and using better tape recordings. However, the information concerning the verbal ability of the field test groups added important additional evidence for explaining the results.

Tables 1 and 2 provide the intercorrelations among listening ability, verbal ability and minority status for the respondents of packages one and

TABLE 1

Intercorrelations Among Listening Ability, Verbal Ability and Minority Status for Package One

	Listening Ability	Verbal Ability	Minority Status
Listening Ability	1.0	.593*	-.086
Verbal Ability		1.0	.052
Minority Status			1.0

TABLE 2

Intercorrelations Among Listening Ability, Verbal Ability and Minority Status for Package Two

	Listening Ability	Verbal Ability	Minority Status
Listening Ability	1.0	.474*	-.232**
Verbal Ability		1.0	-.255**
Minority Status			1.0

*$p < .01$
**$p < .05$

two. It is clear that the relationship between listening ability and verbal ability for both sets of items was high, .59 correlations for package one and .47 for package two. This is in line with the results of previous studies of this relationship (Crook, 1957; Haberland, 1959). The differences in the results of the two packages was indicated by the relationships between the factors of listening ability and verbal ability, and the factor of minority status. For package one, there was no significant correlation between listening ability and minority status, nor between verbal ability and

minority status. For package two, the correlation between listening ability and minority status was -.23 and between verbal ability and minority status was -.26. These correlations were of similar magnitude and indicated significantly different ($P < .05$) responses of minority and
dents. The minority students were performing below the level of the nonminority students on both measures. It should be recalled that the verbal abili[illegible] of the minority students who responded to package one was above their group average, and the verbal ability of the minority students who responded to package two was well below their group average. This collective evidence seems to indicate that the differing performance of minority students on the listening measures may be attributed to differing levels of verbal ability rather than to racial/ethnic bias within the items themselves.

Discussion

The results of the second listening study confirms earlier studies regarding a positive correlation between listening ability and verbal ability. It also suggests that differing responses of minority and nonminority students to listening items in the first study may be attributed to differing levels of verbal ability in the group and not to problems of racial/ethnic bias in the items. In retrospect, this seems a reasonable explanation considering the make up of the first field test sample. In the first study, the nonminority students mostly came from small cities with generally affluent school populations. The minority students mostly came from large cities with generally poor school populations. However, the second study does not fully explain why in the original study the differences in

responses of minority and nonminority students were so much greater with the listening items than with the items which were assessing other communication competencies--informative speaking, persuasive speaking, ritualizing and sharing feelings. There is a suggestion from the two studies that listening ability may be tied more directly to verbal ability than other communication competencies. This finding presents interesting possibilities for further study.

References

Crook, F. E. "Interrelationship Among a Group of Language Arts Tests." Journal of Educational Research, 51 (1957), 305-311.

Haberland, J. A. "A Comparison of Listening Tests with Standardized Tests." Journal of Educational Research, 52 (1959), 299-302.

Harris, M. L. "Some Methodological Suggestions for Construction of an Objective Measuring Instrument." Madison, Wisconsin: Wisconsin Research and Development Center for Cognitive Learning, University of Wisconsin, 1968.

Lorge, I., Thorndike, R. L., and Hagen, E. The Lorge-Thorndike Intellegence Tests. Boston: Houghton Mifflin, 1966.

Mead, N. "The Development of an Instrument for Assessing Functional Communication Competence of Seventeen-Year-Olds." Unpublished doctoral dissertation, University of Denver, Denver, Colorado, 1977(a).

Mead, N. "Issues Related to National Assessment of Speaking and Listening Skills." Unpublished paper presented at the Speech Communication Association Annual Convention, Washington, D.C., 1977(b).

Stanley, J. C. and Hopkins, K. D. Educational and Psychological Measurement and Evaluation. Englewood Cliffs, New Jersey: Prentice-Hall, 1972.

08-SC-53

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
Education Commission of the States
1860 Lincoln St., Suite 700
DENVER, CO 80295

Reprinted 10/80